

(19)



Europäisches Patentamt  
European Patent Office  
Office européen des brevets



(11)

**EP 1 170 678 A2**

(12)

**EUROPÄISCHE PATENTANMELDUNG**

(43) Veröffentlichungstag:  
09.01.2002 Patentblatt 2002/02

(51) Int. Cl. 7: **G06F 17/30**(21) Anmeldenummer: **01114261.9**(22) Anmeldetag: **12.06.2001**

(84) Benannte Vertragsstaaten:  
**AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU  
MC NL PT SE TR**  
Benannte Erstattungsstaaten:  
**AL LT LV MK RO SI**

(71) Anmelder: **PixFind Net Technologies GmbH**  
10587 Berlin (DE)

(72) Erfinder: **Zuse, Horst, Dipl.-Ing.**  
10719 Berlin (DE)

(30) Priorität: **03.07.2000 DE 10032225**

(74) Vertreter: **Liesegang, Eva**  
**Forrester & Boehmert, Pettenkoferstrasse 20-22**  
80336 München (DE)

**(54) Verfahren und Vorrichtung zur automatischen Suche relevanter Bilddatensätze**

(57) Die Erfindung bezieht sich auf ein Verfahren und eine Vorrichtung zur automatischen Suche relevanter Bilddatensätze aus einer Menge von  $n$  ( $n \geq 2$ ) in einer Speichereinrichtung elektronisch gespeicherten Bilddatensätzen, wobei in der Speichereinrichtung für jeden der Endbilddatensätze Bildeigenschaften elektronisch gespeichert sind, und wobei die  $n$  Bilddatensätze die gespeicherten Bildeigenschaften mit Hilfe von Prozessormitteln elektronisch verarbeitbar sind. Eine Trainingsmenge, die die elektronisch gespeicherten Bildeigen-

schaften von Bilddatensätzen umfaßt, welche von einem Benutzer ausgewählt wurden, wird im Rahmen eines mit Hilfe der Prozessormittel automatisch geführten maschinellen Lernverfahrens genutzt, um eine Entscheidungsfunktion  $f$  zu ermitteln. Mit Hilfe der Entscheidungsfunktion  $f$  werden anschließend Systemrelevanzen der Bilddatensätze berechnet. In Abhängigkeit von der jeweiligen Systemrelevanz werden dann ausgewählte Bilddatensätze im Rahmen des Suchergebnisses zur Ausgabe zur Verfügung gestellt.

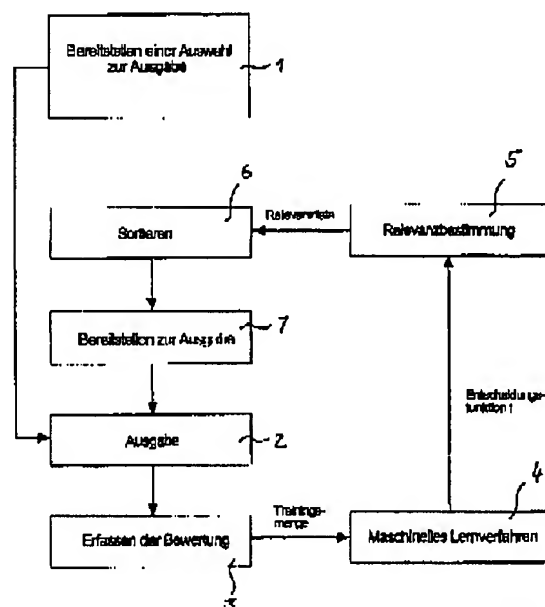


Fig. 1

**EP 1 170 678 A2**

## EP 1 170 676 A2

## Beschreibung

[0001] Die Erfindung liegt auf dem Gebiet der Suche bzw. des Abrufs von Information, insbesondere von elektronisch gespeicherten Bildern in elektronischen Datenbanken.

5 [0002] Im Zusammenhang mit dem Internet, welches in kürzester Zeit immer größere Informationsmengen zur Verfügung stellt, besteht ein wachsender Bedarf für Möglichkeiten zur gezielten und effektiven Suche nach Informationen, die in Datenbanken elektronisch gespeichert sind. Die Problemstellung in Verbindung mit der Suche bzw. dem Abruf von elektronisch gespeicherten Informationen in Datenbanken, welche als sogenanntes "Information Retrieval" bezeichnet wird, kann wie folgt formuliert werden: Gegeben ist eine Datenbank mit  $n$  Datensätzen  $x_1, \dots, x_n$  ( $n \geq 2$ ). Die  
10 Bildsuche bzw. der Abruf von Bildern ist ein spezielles Verfahren der Suche in Datenbanken. Bei der Bildsuche handelt es sich bei den Datensätzen  $x_1, \dots, x_n$  um  $n$  in elektronischer Form vorliegende Bilder. Gesucht ist eine Teilmenge  $D_{rel}$  von relevanten Datensätzen (elektronischen Bildern) aus der Datenbank. Die Teilmenge  $D_{rel}$  stellt die für eine bestimmte Fragestellung eines Benutzers relevante Menge von Datensätzen dar. Im Fall der Bildsuche sind dies beispielsweise Strandbilder von der Küste Hawai's.

15 [0003] Bei bekannten Verfahren und Vorrichtungen zur Bildsuche in Datenbanken wird versucht, die relevante Teilmenge  $D_{rel}$  zunächst mit Hilfe von Schlagwörtern zu beschreiben, und anschließend eine Suchanfrage mit Hilfe der Schlagwörter ausgeführt. Der Datenbankbenutzer stellt hierbei - meist ohne Kenntnis der gesamten in der Datenbank benutzten Schlagwortliste - eine textuelle Anfrage; im oben genannten Beispiel könnte die Anfrage die Worte "Strand Hawai" umfassen. Die textuelle Anfrage wird dann mit Schlagwörtern verglichen, die für die von der Datenbank um-  
20 faßten Bilder jeweils gespeichert sind. Hierbei wird häufig das sogenannte Boolesche Suchverfahren angewendet. Bei diesem Verfahren hat der Benutzer die Möglichkeit, die Schlagwörter mit AND, OR und NOT zu verbinden. In manchen Verfahren bzw. Vorrichtungen ist es darüber hinaus möglich, diese drei Operationen jeweils mit einer Wichtung zu versehen.

26 [0004] Im Zusammenhang mit der Suche in einer Bilddatenbank treten die folgenden Probleme auf:

1. Wie kann die gesuchte Teilmenge  $D_{rel}$  für die Suche mit Hilfe von Worten systematisch beschrieben werden, wenn die Datensätze  $x_i$  (digitalisierte) Bilder sind?
2. Die Datenbank umfassen oft eine sehr große Anzahl von Bildern ( $n \gg 100.000$ ), so daß der Benutzer sich nicht alle  $n$  Bilder anschauen und beurteilen kann.

30 [0005] Beim Beschreiben von Bildern für die Suche kann grundsätzlich zwischen zwei unterschiedlichen Ansätzen unterschieden werden. In einem Fall wird das Bild digitalisiert, und es werden Merkmale aus dem digitalen Bild extrahiert. Dieses beginnt bei der einfachsten Beschreibung mit Hilfe von Grau- oder von Farbwerten jedes Pixels (sogenannten "low-level features"), d.h. bei einem Bild mit  $1.000 \times 1.000$  Pixeln werden 1.000.000 unterschiedliche Merkmale je Bild extrahiert. Es endet bei Merkmalen, die als "high-level features" bezeichnet werden, wie der Anzahl von Kanten und Ecken oder der Anzahl von Flächen etc. Der Vorteil der Verwendung einfacher Merkmale ist deren schnelle Berechenbarkeit; der Nachteil ist allerdings, daß diese Merkmale sehr schlecht geeignet sind, um relevante Suchbild-  
35 mungen zu beschreiben. Obwohl komplexe Merkmale hier deutlich geeigneter wären, ist ihre Extraktion zur Zeit noch mit einem so hohen Aufwand verbunden, daß eine Anwendung im Bereich von Datenbeständen mit mehr als 10.000 Bildern aus praktischen Gesichtspunkten nahezu unmöglich ist.

40 [0006] Bei einem anderen bekannten Verfahren wird ein Bild von einem Menschen "verschlagwortet", d.h. zu jedem Bild wird eine Liste von textuellen Schlagwörtern erstellt, die sich auf den Bildinhalt beziehen. Die Vorteile dieser aufwendigen Merkmalsextraktion bestehen darin, daß eine Charakterisierung von relevanten Bildern in Form von Verknüpfungen der Schlagwörter einfach möglich ist. Aus technischer Sicht wird ein Bild  $x$  mittels eines Vektors  $x \in \{0, 1\}^s$  dargestellt ( $s$  ist die Anzahl aller möglichen Schlagwörter); wenn das  $i$ -te Schlagwort in der Schlagwortliste des  
45 Bildes vorhanden ist, so ist die  $i$ -te Komponente  $x_i$  des Vektors  $x$  gleich 1, andernfalls 0. Verknüpfungen wie Konjunktion (AND) oder Disjunktion (OR) von Worten können in diesem Fall mittels mathematischer Verknüpfungen, wie Multiplikation und Addition dargestellt werden.

50 [0007] Die Bildsuchmaschine berechnet nach dem Start der Suchanfrage für jeden der elektronisch gespeicherten Bilddatensätze  $x_1, \dots, x_n$  eine Systemrelevanz bezüglich der Suchanfrage. Die Berechnung einer jeweiligen Systemrelevanz ist eine wesentliche Eigenschaft jeder Bildsuchmaschine bzw. jedes Bildsuchverfahrens. Die Effektivität und die Qualität der Berechnung der jeweiligen Systemrelevanz ist für den Erfolg des Suchsystems von wesentlicher Bedeutung. Zwei prinzipiell unterschiedliche Ansätze haben sich in Verbindung mit der Berechnung der Systemrelevanz bei Schlagwort-Suchverfahren durchgesetzt:

- Wenn die vom Benutzer erzeugte textuelle Suchanfrage, welche nur Schlagwörter umfaßt, als ein Vektor  $q \in \{0, 1\}^s$  interpretiert wird, ist es möglich, mit Hilfe der für die elektronisch gespeicherten Bilder vorhandenen Schlagwortlisten die Ähnlichkeit der textuellen Suchanfrage mit der jeweiligen Schlagwortliste der Bilder bzw. der Bilda-

## EP 1 170 678 A2

tensätze in der Datenbank zu berechnen und diese Ähnlichkeit als Systemrelevanzmaß zu benutzen. Dieser Ansatz ist als "vector space model" bekannt und beispielsweise in der Druckschrift G. Salton: "Automatic Information Organization and Retrieval", McGraw-Hill, New York, 1968, beschrieben.

- 5 - Bei einem anderen Ansatz ist es mit Hilfe eines Wahrscheinlichkeitsmodells für die Schlagworte in relevanten Dokumenten (geschätzt aus der textuellen Suchanfrage, die nur Schlagworte enthält) möglich, die Wahrscheinlichkeit zu berechnen, daß ein Bild von der Teilmenge  $D_{rel}$  umfaßt ist, und diese Wahrscheinlichkeit als Systemrelevanzmaß zu benutzen.
- 10 **[0008]** Basierend auf den gefundenen Systemrelevanzen für alle Bilder der Bilddatenbank werden die Bilder anschließend in Abhängigkeit von den berechneten Systemrelevanzen geordnet und dem Benutzer angeboten. In der Praxis genügt es oft, nur die 100 Bilder mit den höchsten Werten der Systemrelevanz zu finden - ein Aufgabe, die deutlich schneller zu lösen ist, als eine wesentliche größere Anzahl, beispielsweise 1.000.000 Bilder zu sortieren.
- 15 **[0009]** Sollte der Datenbankbenutzer mit dem Suchergebnis noch nicht zufrieden sein, so muß er zurück zur textuellen Anfrage und diese verändern, z.B. weiter einschränken. In einigen Systemen hat der Benutzer die Möglichkeit, eine Rückkopplung ("feedback") in Form der Auswahl eines Bildes zu geben, welches er für "sehr ähnlich" oder "nahe" zu den relevanten Dokumenten  $D_{rel}$  hält.
- 20 **[0010]** Ein wesentlicher Nachteil, den solche Verfahren aufweisen, ist die Tatsache, daß Suchanfragen auf der Basis identischer textueller Vorgaben des Benutzers in Verbindung mit einem bestimmten Bildbestand stets das gleiche Suchergebnis liefern. Dies bedeutet, daß sich die Benutzer bei der Bildsuche an die Verschlagwortung der Datenbank anpassen müssen, um die individuellen Präferenzen und Eigenarten der Datenbank modellieren zu können, da der einzige Weg, wie Datenbankbenutzer und Suchsystem miteinander "kommunizieren" können, die textuelle Suchanfrage ist. Dieses erfordert von dem Benutzer eine intensive und in der Regel zeitaufwendige "Erkundung" der Spezifik der Verschlagwortung der jeweiligen Datenbank, in welcher der Benutzer die Suche ausführt.
- 25 **[0011]** Aufgabe der Erfindung ist es, ein verbessertes Verfahren und eine verbesserte Vorrichtung für die Suche nach einer relevanten Teilmenge von Datensätzen aus einer Menge von Datensätzen, insbesondere von Bilddatensätzen zu schaffen, die in einer Datenbank elektronisch gespeichert sind, wobei die Effektivität und die Qualität der Suche sowie die Bedienerfreundlichkeit verbessert werden sollen.
- 30 **[0012]** Nach einem Aspekt der Erfindung ist ein Verfahren zur automatischen Suche relevanter Bilddatensätze aus einer Menge von  $n$  ( $n \geq 2$ ) in einer Speichereinrichtung elektronisch gespeicherten Bilddatensätzen geschaffen, wobei in der Speichereinrichtung für jeden der  $n$  Bilddatensätze Bildeigenschaften elektronisch gespeichert sind, und wobei die  $n$  Bilddatensätze und die gespeicherten Bildeigenschaften mit Hilfe von Prozessormitteln elektronisch verarbeitbar sind, das Verfahren die folgenden Verfahrensschritte aufweisend:
  - 35 a) Bereitstellen einer ersten Auswahl von Bilddatensätzen aus den  $n$  Bilddatensätzen mittels der Prozessormittel zur Ausgabe mit Hilfe einer Monitoreinrichtung;
  - b) Ausgeben mehrerer der Bilddatensätze der ersten Auswahl von Bilddatensätzen mit Hilfe der Monitoreinrichtung;
  - 40 c) elektronisches Erfassen einer jeweiligen Bewertung eines Benutzers für wenigstens einen relevanten Bilddatensatz aus den mehreren gemäß Verfahrensschritt b) ausgegebenen Bilddatensätzen; und
  - d) Bereitstellen einer zweiten Auswahl von  $m$  ( $m < n$ ) Bilddatensätzen aus den  $n$  Bilddatensätzen in einer von einer jeweiligen Systemrelevanz der  $m$  Bilddatensätze abhängigen Reihenfolge zur Ausgabe mittels der Monitoreinrichtung;
- 45 wobei zum Bereitstellen der zweiten Auswahl der  $m$  Bilddatensätze ein maschinelles Lernverfahren zum elektronischen Ermitteln einer Entscheidungsfunktion  $f$  ausgeführt wird; wobei die für den wenigstens einen relevanten Bilddatensatz in der Speichereinrichtung elektronisch gespeicherten Bildeigenschaften eine Trainingsmenge für das maschinelle Lernverfahren bilden; wobei für  $k$  ( $k \geq m$ ) Bilddatensätze mit Hilfe der Entscheidungsfunktion und der jeweiligen elektronisch gespeicherten Bildeigenschaften die jeweilige Systemrelevanz ermittelt wird; und wobei die  $k$  Bilddatensätze
- 50 zumindest einen Teil der  $m$  Bilddatensätze der zweiten Auswahl umfassen.
- [0013]** Nach einem weiteren Aspekt der Erfindung ist eine Bildsuchvorrichtung geschaffen, die Bildsuchvorrichtung aufweisend:
  - 55 - einer Speichereinrichtung zum elektronischen Speichern von  $n$  ( $n \geq 2$ ) Bilddatensätzen und jeweiligen Bildeigenschaftsdaten, die den  $n$  Bilddatensätzen jeweils zugeordnet sind;
  - einer Monitoreinrichtung zum Ausgeben einer ersten Auswahl von Bilddatensätzen aus den  $n$  Bilddatensätzen;
  - Erfassungsmitteln zum elektronischen Erfassen einer jeweiligen Bewertung eines Benutzers für wenigstens einen relevanten Bilddatensatz aus der ersten ausgegebenen Auswahl von Bilddatensätzen;

## EP 1 170 678 A2

- Prozessormitteln zum automatischen Ausführen eines maschinellen Lernverfahrens, um eine Entscheidungsfunktion unter Berücksichtigung der für den wenigstens einen relevanten Bilddatensatz in der Speichereinrichtung elektronisch gespeicherten Bildeigenschaften zu ermitteln, und zum elektronischen Ermitteln einer jeweiligen Systemrelevanz für  $m$  ( $m \leq n$ ) Bilddatensätze der  $n$  in der Speichereinrichtung elektronisch gespeicherten Bilddatensätze mit Hilfe der Entscheidungsfunktion  $f$  und der jeweiligen elektronisch gespeicherten Bildeigenschaften; und
  - Bereitstellungsmitteln zum Bereitstellen einer zweiten Auswahl von Bilddatensätzen in einer von der jeweiligen Systemrelevanz abhängigen Reihenfolge zur Ausgabe mittels der Monitoreinrichtung, wobei die zweite Auswahl von Bilddatensätzen  $k$  ( $k \geq m$ ) Bilddatensätze der  $m$  Bilddatensätze umfaßt.
- 10 [0014] Die Erfindung umfaßt den wesentlichen Grundgedanken, eine von dem Benutzer einer Bildsucheinrichtung ausgewählte Teilmenge relevanter Bilder bzw. die zugehörigen Bilddatensätze als eine Trainingsmenge für ein maschinelles Lernverfahren der Bildsucheinrichtung zu nutzen. Im Rahmen des elektronisch ausgeführten maschinellen Lernverfahrens werden die den Bildern aus der Trainingsmenge zugeordneten Bildeigenschaften elektronisch verarbeitet, wobei mathematische Verfahren genutzt werden, um relevante Bilddatensätze in einer Datenbank einer erfaßten
- 15 Suchanfrage entsprechend aufzusuchen und für eine Weiterverarbeitung, insbesondere die Ausgabe über einen Monitor zur Verfügung zu stellen.
- [0015] Ein wesentlicher Vorteil der Erfindung besteht darin, daß die unkomfortable Suche mittels Schlagworten durch einfaches Bewerten von Suchergebnissen, beispielsweise per Mausklick ersetzt werden kann. Auf diese Weise wird der Benutzer des Suchsystems von der im Stand der Technik notwendigen Bildbeschreibung (digitale Merkmale und/
- 20 oder Schlagworte) befreit.
- [0016] Ein weiterer Vorteil der Erfindung besteht darin, daß sich das Suchsystem mit Hilfe der Nutzung des maschinellen Lernverfahrens an den Benutzer anpaßt, und nicht umgekehrt, wie es beim Erlernen der spezifischen Verschlagwortung der Datenbank durch den Benutzer der Fall ist. Während bei den bekannten Suchsystemen davon ausgegangen wird, daß das Bildsuchsystem (intern) mit einer Suchanfrage  $q$  arbeitet, wird dieses Paradigma durch das maschinelle Lernen einer Entscheidungsfunktion  $f$  abgelöst. In der Praxis bedeutet dies, daß kein allgemeingültiges Ähnlichkeitsmaß zwischen Textanfrage  $q$  und Bildern  $x$  gefunden werden muß.
- 25 [0017] Das neue Verfahren kann in Verbindung mit bestehenden elektronisch gespeicherten Bildbeschreibungen genutzt werden und erfordert keine aufwendige neue Vorverarbeitung von Bilddatenbanken. Darüberhinaus ist das Verfahren in der Lage, (einfach zu extrahierende) digitale Merkmale zu benutzen, um die Systemrelevanz bestmöglich der Benutzerrelevanz (ausgedrückt durch die dem System unbekannte Menge  $D_{rel}$ ) anzupassen.
- 30 [0018] Da sich der Benutzer des neuen Suchverfahrens bzw. der neuen Suchvorrichtung bei der wiederholten Ausführung des Verfahrens nach jeder Ausgabe von Suchergebnissen erneut für relevante Bilder entscheiden kann, d.h. eine erneute, von der vorherigen Bewertung unabhängige Bewertung vornehmen kann, wird ein sogenanntes "creative drifting" unterstützt. Dieses bedeutet, daß ein Benutzer, der anfänglich Strandbilder von Hawai sucht, sich bei der erneuten Bewertung anders entscheiden kann. Findet der Benutzer im ausgegebenen Suchergebnis ein Motiv, das seinem Suchwunsch eher entspricht, beispielsweise einen Strand in Australien, so kann er nur noch Bilder von Australiens Stränden als relevant bewerten. Auf diese Weise wird das ursprüngliche Konzept einer Küste Hawai's bei der erneuten Bewertung ignoriert.
- 35 [0019] Eine zweckmäßige Weiterbildung der Erfindung sieht vor, daß die erste Auswahl von Bilddatensätzen aus den  $n$  Bilddatensätzen mittels einer Schlagwortsuche ermittelt wird, wodurch das neue Verfahren auf einfache Weise mit der aus dem Stand der Technik bekannten Schlagwortsuche kombinierbar ist.
- 40 [0020] Eine hinsichtlich einer verbesserten Benutzerfreundlichkeit bevorzugte Fortbildung der Erfindung sieht vor, daß das elektronische Erfassen der jeweiligen Bewertung des Benutzers für den wenigstens einen relevanten Bilddatensatz gemäß Verfahrensschritt c) die Erfassung einer Betätigung eines elektronischen Auswahlmittels, insbesondere einer Mauseinrichtung umfaßt, welches mit der Monitoreinrichtung zusammenwirkt. Insbesondere bei der Verwendung der Mauseinrichtung ist ein Verfahren geschaffen, welches es dem Benutzer ermöglicht, sich bei der Auswahl der Bilddatensätze auf die dargestellten Datensätze zu konzentrieren und gleichzeitig auf einfache Weise eine Auswahl zu treffen.
- 45 [0021] Zur Minimierung des elektronischen Rechenaufwands und zur Beschleunigung des Verfahrens kann bei einer bevorzugten Ausgestaltung der Erfindung vorgesehen sein, daß die jeweilige Bewertung des Benutzers für den wenigstens einen relevanten Bilddatensatz gemäß Verfahrensschritt c) als eine binäre Bewertung elektronisch erfaßt wird, so daß jeder bewertete Bilddatensatz als ein relevanter Bilddatensatz und jeder nicht bewertete Bilddatensatz als ein nicht-relevanter Bilddatensatz erfaßbar ist.
- 50 [0022] Um das Verfahren hinsichtlich der Verfahrensdauer zu minimieren, kann eine zweckmäßige Weiterbildung der Erfindung vorsehen, daß die Entscheidungsfunktion  $f$  im Rahmen des maschinellen Lernverfahrens so ermittelt wird, daß das elektronische Ermitteln der jeweiligen Systemrelevanz hinsichtlich der hierfür benötigten Zeitdauer optimiert ist.
- 55 [0023] Das maschinelle Lernverfahren kann zweckmäßig mit Hilfe eines Perzeptron-Lernverfahrens ausgeführt werden.

## EP 1 170 676 A2

den.

[0024] Um das maschinelle Lernverfahren auch in Verbindung mit vorhandenen Datenbanken ausführen zu können, sieht eine bevorzugte Ausführungsform der Erfindung vor, daß gespeicherten Bildeigenschaften mit Hilfe der Prozessmittel elektronisch auswertbare Schlagworte umfassen.

5 [0025] Die Erfindung wird im Folgenden anhand von Ausführungsbeispielen unter Bezugnahme auf eine Zeichnung näher erläutert.

[0026] Figur 1 zeigt eine schematische Darstellung eines Blockdiagramms zur Beschreibung eines Verfahrens zur automatischen Suche relevanter Bilddatensätze.

10 [0027] Eine zur Suche zur Verfügung stehende Menge  $D$  von Bildern ist in elektronischer Form mit Hilfe von  $n$  Bilddatensätzen  $x_1, \dots, x_n$  ( $n \geq 2$ ) in einer Datenbank gespeichert. Im Folgenden wird davon ausgegangen, daß die Bilder verschlagwortet sind, d.h. für jedes der Bilder sind Bildeigenschaften in Form einer Schlagwortliste gebildet, wobei die jeweilige Schlagwortliste in der Datenbank als elektronisch recherchierbarer Schlagwort-Vektor gespeichert ist. Für die Ausführung des erfindungsgemäßen Verfahrens können jedoch beliebige andere Bildeigenschaften genutzt werden, wenn diese in elektronischer Form speicherbar und verarbeitbar sind. Beispielsweise können digitale Merkmale, Grau- oder Farbwerte oder Farbhistogramme genutzt werden. Besonders die Einbeziehung von Farbhistogrammen würde es erlauben, Atmosphäre, Stimmung und Stil des jeweiligen Bildes zu berücksichtigen. Eine Ausnutzung der Struktur und der Form des Bildes ist beispielsweise mit Hilfe von Merkmalen aus einer Wavelet-Analyse möglich. Die Nutzung von Schlagwortlisten hat den Vorteil, daß das Suchsystem wahlweise mit klassischen Schlagwort-Suchtechniken kombiniert werden kann. Auch eine kombinierte Einbeziehung der Schlagworte und eine oder mehrerer digitaler Merkmale kann zweckmäßig vorgesehen sein, um die Qualität der Suchergebnisse zu verbessern.

20 [0028] In diesem Zusammenhang ist anzumerken, daß die Bildbeschreibung ein entscheidender Faktor für die Qualität jedes Bildsuchsystems darstellt. Wenn die elektronisch gespeicherten Bildeigenschaften (Schlagwortlisten) der Bilder bzw. Bilddatensätze nichts mit den Bildinhalten zu tun haben, so kann man von keinem Bildsuchsystem sinnvolle Ergebnisse erwarten.

25 [0029] Gemäß Figur 1 wird zunächst eine erste Auswahl von Bildern, d.h. der zugehörigen Bilddatensätze zur Ausgabe über eine Monitoreinrichtung (nicht dargestellt) bereitgestellt 1, die mit einem Computer verbunden ist. Der Computer ist seinerseits mit der Datenbank verbunden, in welcher die Menge  $D$  von Bilddatensätzen elektronisch gespeichert ist. Die erste Auswahl kann mit Hilfe einer Schlagwortsuche erzeugt werden. Es können auch andere Suchverfahren genutzt werden. Wenn die Schlagwortsuche genutzt wird, ist sichergestellt, daß das im Folgenden beschriebene Verfahren funktionale Erfordernisse von bekannten Schlagwort-Suchsystemen erfüllt, d.h. mit diesen zusammenwirken kann bzw. in Verbindung mit diesen genutzt werden kann.

30 [0030] Anschließend werden zumindest einige der Bilder der ersten Auswahl mit Hilfe der Monitoreinrichtung ausgegeben 2 (vgl. Figur 1). Der Benutzer kann dann mit Hilfe von Auswahlmitteln des Computers zu mindestens einigen Suchergebnissen (ausgegebenen Bildern bzw. Bilddatensätzen) eine jeweilige (binäre) Wertung  $b_i \in \{-1, +1\}$  abgeben, wobei -1 bedeutet, daß das entsprechende Bild nichts mit der gesuchten, relevanten Teilmenge von Bildern  $D_{rel}$  zu tun hat, während +1 eine gewisse "Nähe" bzw. Übereinstimmung des jeweiligen Bildes zu den relevanten Bildern  $D_{rel}$  ausdrückt. Hierbei ist es nicht notwendig, daß alle präsentierten bzw. ausgegebenen Bilder bewertet werden. Die Bewertung führt der Benutzer beispielsweise mittels eines einfachen Mausklicks aus, welcher mit Hilfe des Computers elektronisch erfaßt wird 3 (vgl. Figur 1) und dem jeweiligen Bild elektronisch zugeordnet wird. Für die Bewertung können beliebige Auswahlmittel des Computers genutzt werden, die die Erfassung der Bewertung derart ermöglichen, daß die erfaßte Bewertung anschließend dem jeweiligen Bild zugeordnet werden kann. Zu den nutzbaren Auswahlmitteln gehört zum Beispiel eine Tastatur des Computers.

35 [0031] In diesem Zusammenhang wird ein wesentlicher Vorteil Verfahrens deutlich. Das Bewerten der ausgegebenen Bilder ist in sehr kurzer Zeit, beispielsweise in weniger als einer Sekunde möglich. Im Gegensatz dazu erfordert die Bewertung eines Textes, wie es bei bekannten Schlagwort-Suchmaschinen notwendig ist, üblicherweise mehrere Minuten. Es ist realistisch anzunehmen, daß der Benutzer in der Regel nicht mehr als 100 Bilder bewerten wird.

40 [0032] Die bewerteten Bilder bzw. Bilddatensätze - welche meist nur einen sehr kleinen Teil der gesamten, für die Suche zur Verfügung stehenden Menge  $D$  ausmachen - bilden eine Trainingsmenge für ein anschließendes maschinelles Lernverfahren 4 (vgl. Figur 1), welches mit Hilfe des Computers automatisch ausgeführt wird. Im Rahmen des maschinellen Lernverfahrens wird versucht, eine reellwertige Entscheidungsfunktion  $f: D \rightarrow \mathbb{R}$  zu lernen, die möglichst genau die Bewertung der vom Benutzer direkt zur Verfügung gestellten bzw. bewerteten Bilder (Trainingsmenge) nachbildet. Hierbei ist die Anwendung von Verfahren wie des als solchen bekannten Perzeptonlernens (M. Rosenblatt: Principles of neurodynamics: Perceptron and Theory of Brain Mechanism, Spartan-Books, Washington D.C., 1962) oder der sogenannten "Support Vector Machine" (C. Cortes und V. Vapnik: Support Vector Networks in Machine Learning, 20:273-297, 1995) möglich. Bei der automatischen elektronischen Ausführung des maschinellen Lernverfahrens wird zweckmäßig die Tatsache ausgenutzt, daß es im Gegensatz zu klassischen Fragestellungen maschineller Lernverfahren nicht von wesentlicher Bedeutung ist, daß ein Vorhersagefehler (der Relevanz) auf nicht zum maschinellen Lernen genutzte Bilder der Menge  $D$  sehr klein ist, da es bei der praktischen Umsetzung des Verfahrens genügt,

## EP 1 170 676 A2

wenn mindestens ein Bild der Teilmenge der relevanten Bilder  $D_{rel}$  gefunden wurde.

[0033] Die im Rahmen des maschinellen Lernverfahrens ermittelte Entscheidungsfunktion  $f$  wird anschließend benutzt, um den Bildern der Menge  $D$  eine neue Relevanz zuzuweisen 5 (vgl. Figur 1), welche zu der vom Benutzer individuell vergebenen Bewertung passend ist. Danach werden die Bilder der neuen Relevanz entsprechend sortiert 6 und zur Ausgabe mittels der Monitoreinrichtung bereitgestellt bzw. angeboten 7 (vgl. Figur 1). Bei der Berechnung der neuen Relevanz, kann es vorteilhaft sein, nur für eine Teilmenge ( $> 10.000$ ) der Gesamtmenge  $D$  von Bildern die neue Relevanz zu berechnen, wenn die der Bildbestand in der Menge  $D$  sehr groß ist (beispielsweise  $> 10.000.000$ ). Hierdurch wird der Rechenaufwand vermindert, was insbesondere zur Verkürzung des hiermit verbundenen Zeitaufwands führt. Zweckmäßig wird die neue Relevanz mit Hilfe der Entscheidungsfunktion hierbei zumindest für die Bilder 10 berechnet, die von der ersten Auswahl 1 (vgl. Figur 1) umfaßt sind.

[0034] Der beschriebene und in Figur 1 schematisch dargestellte Ablauf kann wiederholt ausgeführt werden, bis der Benutzer ein oder mehrere relevante(s) Bild(er) gefunden hat oder eine vollständig neue Suchanfrage starten möchte, d.h. die Menge  $D_{rel}$  sich ändert.

[0035] Eine Kernkomponente des beschriebenen Verfahrens ist die Verwendung des maschinellen Lernverfahrens, da auf diese Weise in einem sogenannten "Information Retrieval System" erstmalig Methoden des Klassifikationslernens mit der Suche von Bildern in elektronischen Datenbanken verbunden werden. Ein maschinelles Lernverfahren ist eine Prozedur, welche aus einer gegebenen Menge von Entscheidungsfunktionen auf der Grundlage der Trainingsmenge eine Entscheidungsfunktion  $f$  auswählt (erlernt). Ziel der Auswahl ist es, diejenige Entscheidungsfunktion  $f^*$  zu finden, welche auf noch nicht gesehenen Bildern der Menge  $D$  von Bildern bzw. Bilddatensätzen die wenigsten Fehlentscheidungen (bezüglich der Relevanz) trifft. Die noch nicht gesehenen Bilder umfassen hierbei die Bilder der Menge  $D$ , welche dem Benutzer für eine Bewertung noch nicht über die Monitoreinrichtung ausgegeben wurden. Der Unterschied zwischen dem Fehler, welchen die gelernte Funktion  $f$  und die beste Funktion  $f^*$  auf die noch nicht gesehenen Bilder erzeugen, wird Generalisierungsfehler genannt. Grundsätzlich lassen sich Lernverfahren nach zwei Kriterien bewerten: 25

- i) Wie klein ist der Generalisierungsfehler im Durchschnitt, d.h. gemittelt über häufiges Lernen?
- ii) Wie aufwendig ist das Lernverfahren in Abhängigkeit von der Anzahl der von der Trainingsmenge umfaßten Trainingsbeispiele ?

[0036] Obwohl es bei vielen Anwendungen darauf ankommt, daß sowohl der Generalisierungsfehler als auch der Rechenaufwand klein sind - ein oft nicht zu realisierendes Ziel - ist es im Falle des hier beschriebenen Suchverfahrens eher von sekundärer Bedeutung, wie aufwendig der Lernalgorithmus ist, da in der Regel maximal 100 vom Benutzer bewertete Bilder im Rahmen der Trainingsmenge zur Verfügung gestellt werden. Es genügt dem Benutzer das Bildsuchsystem oft, wenn er bereits eines der relevanten Bilder innerhalb der ersten 30 Bilder präsentiert bekommt.

[0037] Grundsätzlich ist in Verbindung mit dem beschriebenen neuen Suchverfahren jedes maschinelle Lernverfahren einsetzbar, welches Vektoren  $x$  (die Bildbeschreibungen, beispielsweise Schlagwortvektoren) als Eingabe hat und reelle Zahlen (die berechneten Systemrelevanzen) als Ausgabe liefert. Allerdings ist es vorteilhaft, wenn das Lernverfahren explizit versucht, Entscheidungsfunktionen zu finden, bei denen die Berechnung der Relevanz auf noch nicht gesehenen Bildern sehr schnell geschieht. Gegenwärtig werden vereinfachte lineare Funktionen bevorzugt: 35

$$f(x) = \sum_{i=1}^N w_i \phi_i(x) \quad (1)$$

wobei die Funktionen  $\phi_i : D \rightarrow \mathcal{R}$  als Basisfunktionen bezeichnet werden.  $w_1, \dots, w_N$  sind die vom Lernverfahren zu bestimmenden Größen. Es kann gezeigt werden, daß es für eine sehr große Klasse von Lernverfahren möglich ist, den folgenden Ansatz zu machen: 45

$$w_i = \sum_{j=1}^l \alpha_j \phi_i(x_j) \quad (2)$$

wobei  $x_1, \dots, x_l$  ( $l \leq n$ ) die vom Benutzer bewerteten Bilder (Bildbeschreibungen) sein sollen. Der Vorteil eines solchen Ansatzes wird ersichtlich, wenn man Gleichung (2) in Gleichung (1) einsetzt: 50

## EP 1 170 678 A2

$$f(x) = \sum_{i=1}^N \left( \sum_{j=1}^l \alpha_j \phi_j(x_j) \right) \phi_i(x) \quad (3)$$

$$= \sum_{j=1}^l \alpha_j \underbrace{\left( \sum_{i=1}^N \phi_i(x_j) \phi_i(x) \right)}_{K(x_j, x)} \quad (4)$$

[0038] Anstelle N Parameter  $w_1, \dots, w_N$  lernen zu müssen, was oft zu Problemen, wie der Überanpassung an die Trainingsmenge (sogenanntes "overfitting") führen kann, sind nur noch  $l < 100$  Parameter  $\alpha_1, \dots, \alpha_l$  zu lernen. Insbesondere im Bereich der Bildsuche ist die Anzahl N der Schlagworte weit größer als die Anzahl bewerteter Bilder l. Weiterhin ergibt sich, daß es ausreicht, die Funktion K, die als Kern bezeichnet wird (vgl. G. Wahba: *Spline Models for Observational Data*, Society for Industrial and Applied Mathematics, Philadelphia, 1990), zu spezifizieren. Bei der praktischen Anwendung wird gegenwärtig der folgende Kern bevorzugt:

$$K(x, x') = \left( \sum_{i=1}^S x_i \cdot x'_i + c \right)^p, \quad c \in \mathbb{R}^+, p \in \{1, 2, \dots\} \quad (5)$$

[0039] Dieser Kern hat den folgenden Vorteil: Wenn  $p=2$  gilt, dann bedeutet dies, daß Basisfunktionen  $\phi$  benutzt werden, die von der Gestalt  $x_i \cdot x_j$  sind, d.h. bei Bildbeschreibungsvektoren  $x \in \{0, 1\}^S$  einer logischen Verknüpfung mittels Konjunktion (AND) des i-ten und j-ten Schlagwortes entsprechen. Für größere  $p$  entspricht dies beispielsweise Merkmalen, wie "i-tes, j-tes und k-tes Schlagwort sind gleichzeitig aufgetreten". Verallgemeinert ausgedrückt bedeutet dies, daß Korrelationen zwischen Schlagworten bis zur p-ten Ordnung erfaßt werden. Die Zahl  $c$  steuert die a-priori Gewichtung von Schlagwortkombinationen.

[0040] Basierend auf der hier offenbarten Beschreibung von Entscheidungsfunktionen sind beispielsweise Anwendungen von Lernverfahren wie (a) "Support Vector Machine" (vgl. beispielsweise C. Cortes und V. Vapnik: *Support Vector Networks in Machine Learning*, 20:273-297, 1995), (b) "Bayes Point Machine" (vgl. beispielsweise R. Herbrich et al.: *Bayes point machines: Bayes point in kernel space*, Proceedings of IJCAI Workshop Support Vector Machines, 1999), (c) "Linear Programming Machine" (vgl. beispielsweise T. Graepel et al.: *Classification on proximity data with LP-machines*, Proceedings of the Ninth International Conference on Artificial Neural Networks, 1999) oder (d) das bekannte Perzeptron (vgl. beispielsweise M. Rosenblatt: *Principles of neurodynamics: Perceptron and Theory of Brain Mechanism*, Spartan-Books, Washington D.C., 1962) nutzbar. Gegenwärtig wird das Perzeptronlernverfahren bevorzugt, da gezeigt werden kann, daß dieses Lernverfahren mit dem geringsten Rechenaufwand garantiert zu Lösungen  $\alpha = (\alpha_1, \dots, \alpha_l)$  führt, bei denen nur wenige Koeffizienten  $\alpha_i$  von Null verschieden sind - eine Tatsache, welche die Berechnung neuer Systemrelevanzen durch die gelernte Entscheidungsfunktion  $f$  sehr beschleunigt. Des weiteren wurde der originale Algorithmus um eine automatische Adaption des sogenannten Ridgeterms erweitert, d.h. wenn das Lernverfahren nach einigen, beispielsweise nach zehn Iterationen durch die gesamte Trainingsmenge noch nicht beendet ist, wird die Kernfunktion verändert zu



## EP 1 170 678 A2

$$K(x, x') = \begin{cases} K(x, x') & \text{wenn } x \neq x' \\ K(x, x') + \lambda & \text{wenn } x = x' \end{cases}$$

[0041] Der Parameter  $\lambda$  wird nun nach und nach erhöht und es kann gezeigt werden, daß dies zur Termination des Algorithmus bzw. Verfahrens führen muß.

[0042] Das beschriebene Verfahren zum automatischen Suchen relevanter Bilddatensätze kann mit Hilfe eines Computerprogramms auf einem digitalen Computer, beispielsweise einem Personal-Computer, einem Netzwerk-Computer oder einer Servereinrichtung installiert und ausgeführt werden. Die vorliegende Anmeldung bezieht sich auch auf ein Computerprogramm-Produkt, welches in eine Speichereinrichtung eines digitalen Computers geladen werden kann, wobei das Computerprogramm-Produkt Softwarecode-Abschnitte zum Ausführen der insbesondere in einem der Ansprüche 1 bis 7 offenbarten Verfahrensschritte umfaßt, wenn das Computerprogramm-Produkt auf dem digitalen Computer genutzt wird.

[0043] Die in der vorstehenden Beschreibung, den Ansprüchen und der Zeichnung offenbarten Merkmale der Erfindung können sowohl einzeln als auch in beliebiger Kombination für die Verwirklichung der Erfindung in ihren verschiedenen Ausführungsformen von Bedeutung sein.

## Patentansprüche

1. Verfahren zur automatischen Suche relevanter Bilddatensätze aus einer Menge von  $n$  ( $n \geq 2$ ) in einer Speichereinrichtung elektronisch gespeicherten Bilddatensätzen, wobei in der Speichereinrichtung für jeden der  $n$  Bilddatensätze Bildeigenschaften elektronisch gespeichert sind, und wobei die  $n$  Bilddatensätze und die gespeicherten Bildeigenschaften mit Hilfe von Prozessormitteln elektronisch verarbeitbar sind, das Verfahren die folgenden Verfahrensschritte aufweisend:

- a) Bereitstellen einer ersten Auswahl von Bilddatensätzen aus den  $n$  Bilddatensätzen mittels der Prozessormittel zur Ausgabe mit Hilfe einer Monitoreinrichtung;
- b) Ausgeben mehrerer der Bilddatensätze der ersten Auswahl von Bilddatensätzen mit Hilfe der Monitoreinrichtung;
- c) elektronisches Erfassen einer jeweiligen Bewertung eines Benutzers für wenigstens einen relevanten Bilddatensatz aus den mehreren gemäß Verfahrensschritt b) ausgegebenen Bilddatensätzen; und
- d) Bereitstellen einer zweiten Auswahl von  $m$  ( $m \leq n$ ) Bilddatensätzen aus den  $n$  Bilddatensätzen in einer von einer jeweiligen Systemrelevanz der  $m$  Bilddatensätze abhängigen Reihenfolge zur Ausgabe mittels der Monitoreinrichtung;

wobei zum Bereitstellen der zweiten Auswahl der  $m$  Bilddatensätze ein maschinelles Lernverfahren zum elektronischen Ermitteln einer Entscheidungsfunktion  $f$  ausgeführt wird; wobei die für den wenigstens einen relevanten Bilddatensatz in der Speichereinrichtung elektronisch gespeicherten Bildeigenschaften eine Trainingsmenge für das maschinelle Lernverfahren bilden; wobei für  $k$  ( $k \geq m$ ) Bilddatensätze mit Hilfe der Entscheidungsfunktion  $f$  und der jeweiligen elektronisch gespeicherten Bildeigenschaften die jeweilige Systemrelevanz ermittelt wird; und wobei die  $k$  Bilddatensätze zumindest einen Teil der  $m$  Bilddatensätze der zweiten Auswahl umfassen.

2. Verfahren nach Anspruch 1, dadurch gekennzeichnet, daß die erste Auswahl von Bilddatensätzen aus den  $n$  Bilddatensätzen mittels einer Schlagwortsuche ermittelt wird.

3. Verfahren nach Anspruch 1 oder 2, dadurch gekennzeichnet, daß das elektronische Erfassen der jeweiligen Bewertung des Benutzers für den wenigstens einen relevanten Bilddatensatz gemäß Verfahrensschritt c) die Erfassung einer Betätigung eines elektronischen Auswahlmittels, insbesondere einer Mauseinrichtung umfaßt, welches mit der Monitoreinrichtung zusammenwirkt.

4. Verfahren nach einem der vorangehenden Ansprüche, dadurch gekennzeichnet, daß die jeweilige Bewertung des Benutzers für den wenigstens einen relevanten Bilddatensatz gemäß Verfahrensschritt c) als eine binäre Be-



**EP 1 170 676 A2**

wertung elektronisch erfaßt wird, so daß jeder bewertete Bilddatensatz als ein relevanter Bilddatensatz und jeder nicht bewertete Bilddatensatz als ein nicht-relevanter Bilddatensatz erfaßbar ist.

- 5 6. Verfahren nach einem der vorangehenden Ansprüche, dadurch gekennzeichnet, daß die Entscheidungsfunktion  $f$  im Rahmen des maschinellen Lernverfahrens so ermittelt wird, daß das elektronische Ermitteln der jeweiligen Systemrelevanz hinsichtlich der hierfür benötigten Zeitdauer optimiert ist.
- 10 6. Verfahren nach einem der vorangehenden Ansprüche, dadurch gekennzeichnet, daß das maschinelle Lernverfahren als ein Perzeptron-Lernverfahren ausgeführt wird.
7. Verfahren nach einem der vorangehenden Ansprüche, dadurch gekennzeichnet, daß die gespeicherten Bildeigenschaften mit Hilfe der Prozessormittel elektronisch auswertbare Schlagworte umfassen.
- 15 8. Bildsuchvorrichtung mit:
  - einer Speichereinrichtung zum elektronischen Speichern von  $n$  ( $n > 2$ ) Bilddatensätzen und jeweiligen Bildeigenschaftendaten, die den  $n$  Bilddatensätzen jeweils zugeordnet sind;
  - einer Monitoreinrichtung zum Ausgeben einer ersten Auswahl von Bilddatensätzen aus den  $n$  Bilddatensätzen;
  - 20 - Erfassungsmitteln zum elektronischen Erfassen einer jeweiligen Bewertung eines Benutzers für wenigstens einen relevanten Bilddatensatz aus der ersten ausgegebenen Auswahl von Bilddatensätzen;
  - Prozessormitteln zum automatischen Ausführen eines maschinellen Lernverfahrens, um eine Entscheidungsfunktion unter Berücksichtigung der für den wenigstens einen relevanten Bilddatensatz in der Speichereinrichtung elektronisch gespeicherten Bildeigenschaften zu ermitteln, und zum elektronischen Ermitteln einer jeweiligen Systemrelevanz für  $m$  ( $m < n$ ) Bilddatensätze der  $n$  in der Speichereinrichtung elektronisch gespeicherten Bilddatensätze mit Hilfe der Entscheidungsfunktion  $f$  und der jeweiligen elektronisch gespeicherten Bildeigenschaften; und
  - 25 - Bereitstellungsmitteln zum Bereitstellen einer zweiten Auswahl von Bilddatensätzen in einer von der jeweiligen Systemrelevanz abhängigen Reihenfolge zur Ausgabe mittels der Monitoreinrichtung, wobei die zweite Auswahl von Bilddatensätzen  $k$  ( $k \geq m$ ) Bilddatensätze der  $m$  Bilddatensätze umfaßt.
- 30 9. Bildsuchvorrichtung nach Anspruch 8, gekennzeichnet durch Suchmittel zum Ausführen einer Schlagwortsuche zum Suchen der ersten Auswahl von Bilddatensätzen.
- 35 10. Computerprogrammprodukt, welches in einen internen Speicher einer Computereinrichtung ladbar ist und welches Computerprogrammabschnitte zum Ausführen der Verfahrensschritte nach einem der Ansprüche 1 bis 7 beim Abarbeiten des Computerprogrammprodukts in der Computereinrichtung umfaßt.
- 40 11. Computerprogrammprodukt nach Anspruch 10, wobei wenigstens die Computerprogrammabschnitte auf einem von der Computereinrichtung lesbaren Speichermedium gespeichert sind.
- 45
- 50
- 55

## EP 1 170 678 A2

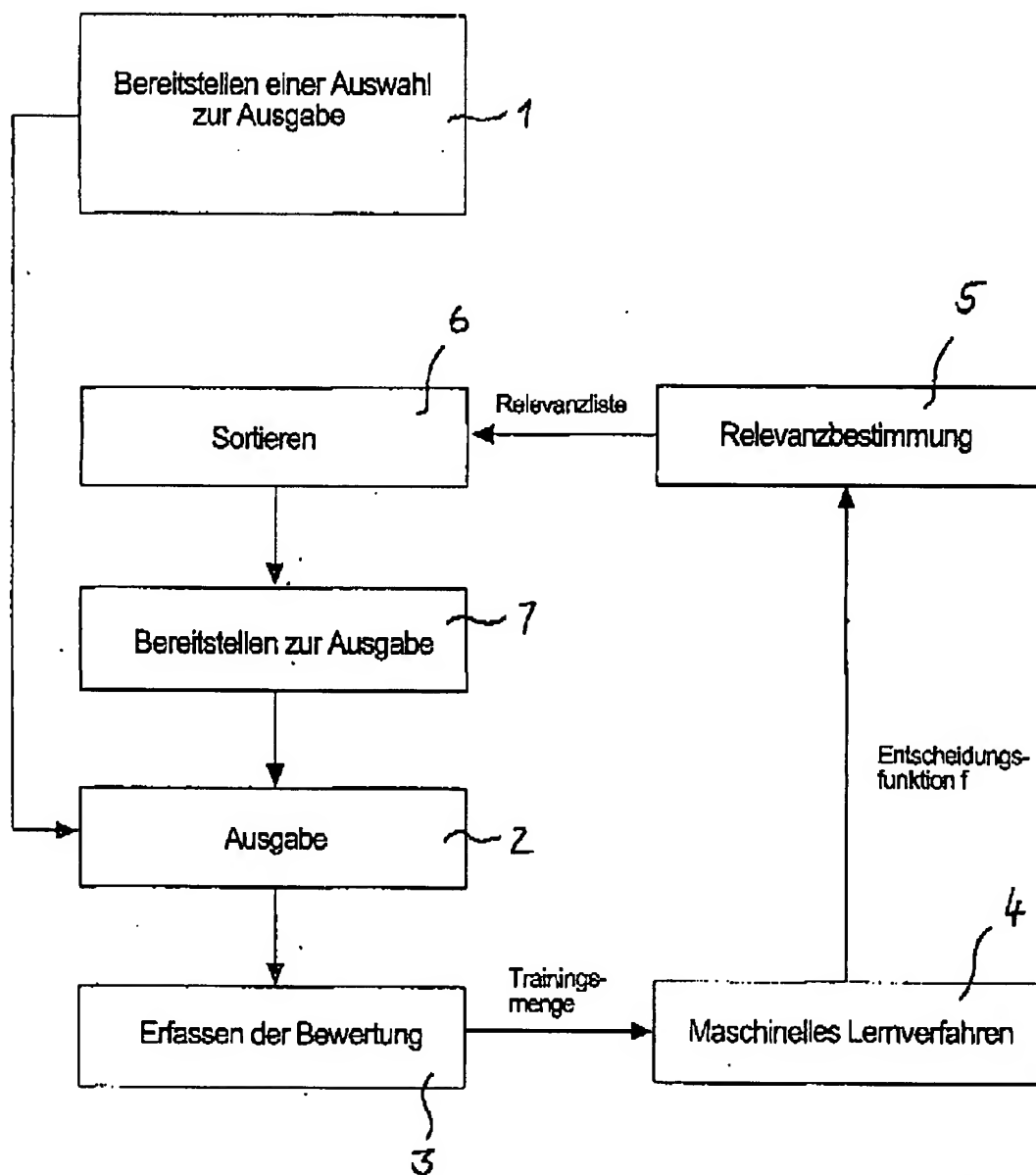


Fig. 1